DOCUMENT RESUME

ED 166 254 TM 008 320

AUTHOR Patterson, Henry O.; Milakofsky, Louis
TITLE A Paper-Pencil Inventory for the Assessment of Piaget's Tasks.
PUB DATE 28 Aug 78
NOTE 47p.; Paper presented at the Annual Meeting of the American Psychological Association (86th, Toronto, Ontario, Canada, August 28, 1978)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Age Differences; Cognitive Development; *Cognitive Tests; *Developmental Stages; Elementary Secondary Education; Higher Education; Multiple Choice Tests; Test Reliability; Test Validity; *Visual Measures
IDENTIFIERS *Inventory of Piagets Developmental Tasks; Paper and Pencil Tests; *Piagetian Tests

ABSTRACT

A project to evaluate Furth's Inventory of Piaget's Developmental Tasks is described, particularly as that test might be useful in adapting instruction and curricula to the cognitive level of science students. This multiple-choice paper-pencil inventory has 72 items in five problem areas: conservation, images, relations, classification, and laws. The test was designed to measure both concrete and formal operations for subjects eight years of age and older. This study measured the reliability of the test for groups in third grade, sixth grade, ninth grade, and college. The lowest reliabilities were found in sixth grade, the age closest to the transition from concrete to formal operations. Test validity was studied several ways. Group and individual administrations were found to give similar results. Age trends followed expected patterns. Correlations with test scores and academic grades ranged from .42 to .87. It was concluded that this is a useful test in certain areas, and that it would be valuable to develop norms. The data were not conclusive, but it appeared that this test would also be useful with mildly or moderately retarded adults. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED166254

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

A Paper-Pencil Inventory for the Assessment of Piaget's Tasks

Henry O. Patterson

Department of Psychology

Louis Milakofsky

Department of Chemistry

The Pennsylvania State University

The Berks Campus

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Henry O. Patterson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

IM008 320

Running Head: A Paper-Pencil Inventory

A Paper-Pencil Inventory for the Assessment of Piaget's Tasks

The need for an easily administered, standardized instrument to measure cognitive development in children and adults has been widely discussed in the educational and psychological literature. Although many Piagetians see attempts at psychometrizing the tasks as invalid and useless (e.g. Elkind, 1971 and Herron, 1978), others have advocated the development of objective, standardized instruments (e.g. Bart, 1978; Kaufman, 1971; Tuddenham, 1971).

The research and applied uses of a standardized instrument have been enumerated well by Tuddenham (1971): (a) as a pretest and posttest to determine the effectiveness of Piaget-based instruction or curriculum; (b) as a means of assessing readiness for specific educational experiences; (c) as a culturally unbiased research tool to investigate age, sex, and social class differences in cognitive development; and (d) as an instrument to empirically investigate Piaget's theory, especially the nuances of the transition from concrete to formal operations, and the development of formal operational thought throughout adulthood. In addition to these considerations, from a purely practical perspective, a paper-pencil test could reduce the arduous, expensive, tedious procedure involved in the administration and scoring of the traditional Piaget interviews for large samples. A valid, streamlined, assessment procedure would expedite much Piaget research and make possible more comprehensive studies.

Already researchers in the area of science education have seen the need for a good assessment tool. A growing number of chemists, physicists, and mathematicians are involved in applied research designed to investigate the implications of recent findings that about 50% of college freshmen enrolled in science courses cannot use formal reasoning (McKinnon & Renner, 1971). Some of these researchers, not well versed in test construction, standardization, or Piaget's theories, and needing a quick assessment instrument suitable for use with groups of students, have reacted to the dearth of suitable tests by devising their own instruments. Although leading science educators such as Arons (1976), Fuller, Karplus, and Lawson (1977), Herron (1975), and Renner and Lawson (1973) have forcefully shown the relevance of Piaget's theories to effective curricula and instruction in science, certain cautions need to be stressed. The danger, as pointed out by Herron (1976, 1977, 1978), with some of these applied research studies by 'naive' but well-intentioned science teachers is that curricula will be prematurely altered and students unfairly categorized on the basis of self-styled Piaget assessment devices which have questionable reliability or validity.

Notwithstanding the theoretical discussions of the usefulness and validity of paper-pencil Piaget tests, and the necessity of their proper standardization, a literature review showed 17 paper-pencil tests have been reported (Barnes, 1977; Bart, 1972; Burney, 1974; Good, Mellon, & Kromhout, 1977; Lawson, 1977; Longeot, 1962;

Lunzer, 1965; McGrew, Wonkka, & Loudin, 1977; Nelson, 1969; Raven, 1973; Renner, 1977; Rowell & Hoffman, 1975; Shayer & Wharry, 1975; Sheehan, 1970; Sills, 1977; Tisher, 1971; and Tomlinson-Keasey, 1975). Almost all of these tests were designed to measure formal thought, most require a high level of verbal and reading ability, and all can be administered in groups. Although most are based on tasks identified by Inhelder & Piaget (1958), different tests use vastly disparate numbers of items and scoring techniques. Most importantly, with few exceptions (e.g. Lawson, 1977 and Sheehan, 1970), very little has been published about both the reliability and validity of these tests.

As well as paper-pencil instruments, the literature includes other notable attempts to standardize batteries of Piaget's tasks which are administered through individual interviews: Goldschmid and Bentler (1968) have standardized and published the Concept Assessment Kit - Conservation; Kaufman (1971) administered a standardized Piaget battery to kindergarten children; and Laurendeau and Pinard (1962) devised an experimental questionnaire to assess causal thinking.

If a global conclusion can be drawn from all the above studies which have taken a psychometric approach to Piaget's tasks, it would be that it is possible to construct assessment instruments which faithfully measure cognitive development and yield results similar to the classic Piaget individual interviews (with the possible exception of the timing for formal thought), and which

greatly facilitate Piaget research. Of all the varied instruments reported, however, there appear to be some notable gaps - no test clearly meets all of the following criteria: (1) comprehensively covers concrete and formal tasks; (2) requires minimal language and reading skills; (3) can be administered to a wide range of ages; (4) can be quickly and easily administered to virtually any size group; (5) can be quickly and objectively scored; and (6) has been standardized and adequately studied for reliability and validity.

A published but little known instrument does exist which seems to meet all but the last criterion. An Inventory of Piaget's Developmental Tasks[1] (IPDT), published by Furth (1970) with B. Ross and J. Youniss at Catholic University's Boys Town Center, is an experimental instrument to be used in the study of cognitive development with subjects eight years and older. Although eight years since its publication, the existence of this inventory has not been widely reported, and there still are no major published reports on reliability or validity.

Although Ross (1968) and Filer (1972) have collected some data on the IPDT which show changes over age, a study by Patterson (1975) was probably the first to use the instrument extensively in a research project. As part of a larger study, he individually administered 40 items from the inventory to 32 5-6-year-olds and 8-9-year-olds and found that the older group scored significantly higher. The study offered no conclusive, empirical evidence of reliability or validity, but concluded that the IPDT had potential

as an instrument to assess cognitive development.

Based on these preliminary indications that the inventory had merit, a major research project was conducted to evaluate the reliability, validity, and usefulness of the IPDT. The initial results of this project have recently been reported (Milakofsky & Patterson, 1977, 1979, in press; Patterson & Milakofsky, 1978).

The purpose of this paper is to expand on these previous reports by presenting additional data on the reliability, validity, and usefulness of the IPDT, and by making suggestions for its improvement as an educational assessment instrument.

Method

Subjects

A total of 542 subjects were tested in all during 1976 and 1977: 250 third, sixth, and ninth grade students in a suburban school district; 210 freshmen and sophomores at The Berks Campus of Penn State University; 60 adults in college continuing education classes at Berks; eight educably retarded children at a special education center; and 14 educably retarded adults at a sheltered workshop. Informed consent was obtained for all participants in the study.

For the major reliability-validity part of the study, approximately four groups of 60 students each were used in the design shown in Table 1. In the public schools, classes were selected by

---

Insert Table 1 about here

---

the respective school principals to participate in the study and individuals were assigned to a group randomly. College students in psychology courses were randomly assigned to one of the three groups. Subjects not in the 12 groups used for the reliability-validity study were adults in selected continuing education psychology courses, additional college freshmen and sophomores in selected chemistry and logic courses, and additional sixth graders from a class in the same school as groups 4-6. The retarded subjects were chosen on availability by the special education teacher in the center and an administrator in the sheltered workshop.

The following background information, where available, was obtained on subjects: age, sex, number of brothers and sisters, birth order, home neighborhood, father's occupation, latest IQ and achievement test scores (SAT scores for college students), and grade averages.

IPDT[2]

The IPDT is a 72-item, untimed, multiple choice, paper-pencil inventory designed originally as part of a project involving the testing of cognitive development in Navaho Indian children. A sample page showing an item in the conservation of volume subtest is included as Figure 1. It was designed to translate some of

---

Insert Figure 1 about here

Piaget's concrete and formal operational tasks into an objective, quick, standardized paper-pencil format which requires minimal

reading ability. Eighteen sets of tasks, divided into subtests of four items each, are presented in illustrations. Each set is introduced by an example which is correctly answered in the test booklet. Five problem areas are covered in the test: conservation (4 subtests), images (4 subtests), relations (3 subtests), classification (4 subtests), and laws (3 subtests). The 18 subtests are identified in Table 2 with the approximate ages for their mastery according to the Piaget literature.

Insert Table 2 about here

In this study, the IPDT was administered in four forms:

1. Group Test. Groups of students were given answer sheets, the IPDT test booklet, and were read instructions on how to take the inventory. Subjects were allowed as much time as necessary to complete all items. Approximately 45 minutes were required for this administration.

2. Group Test Part 1. The same procedure was followed as with the group test except only half the test was given, i.e. subjects only answered items 2 and 3 in each subtest. Items 1 and 4 in each subtest were covered in the booklet and on the answer sheet. This administration usually required 25 minutes.

3. Group Test Part 2. The same as group test part 1 except that subjects answered the first and fourth items in each subtest, items 2 and 3 were covered. Aproximately 25 minutes was required.

4. Individual Test. To determine if the paper-pencil format obscured subjects' understanding of the concepts tested, the second and third items in each subtest (same items as in group test part 1) were chosen to present individually to subjects. Objects similar or identical to those illustrated in the IPDT booklet were collected or constructed and spread out on tables. Subjects were brought in individually, shown the objects and asked questions similar to those in the booklet. Testers were trained not to ask leading questions, but to make sure subjects understood what was being asked. Testers recorded responses on an answer sheet and wrote comments when appropriate. This administration required about 30 minutes.

Testing Procedure

As Table 1 indicates, for the reliability-validity study there were three phases of testing with at least a month elapsing between each phase. In phase one students in each group took tests in one of three orders: 1) one-third took an individual test first, followed by the group test a few days later (groups 1, 4, 7 and 10); 2) a second third had the group test first followed by the individual test (groups 2, 5, 8 and 11); and a final third had only group test part 1 (groups 3, 6, 9 and 12).

In the second phase, the first two groups in each grade took retests of the group test, and the third group took the group test part 2. The third phase consisted of tape-recorded interviews with 30 randomly chosen students in groups 3, 6 and 9; they were asked

to explain their answers to selected questions on which their group had performed poorly. The omission of group 12 interviews was simply due to lack of time before the subjects finished school.

The above counterbalanced design was used so that reliability and validity information could be assessed without contamination by prior exposure to the inventory. The testing sequence within each grade permitted uncontaminated group scores to be compared with uncontaminated individual scores, and test-retest correlations could be computed without administration order contamination. Only groups 3, 6 and 9 were interviewed because all other groups had been contaminated by the individual test.

The subjects who were not part of the above design were administered the IPDT group test; the retarded subjects had to be given special individual administrations where items were read to them out of the booklet.

Results and Discussion

Reliability

Three measures of reliability were computed:

1. Test-retest reliability (coefficient of stability) was determined by correlating total IPDT group test scores for eight groups on two administrations of the group test one month apart. The results are shown in Table 3. These correlations, with one exception, are highly significant and are in the general range of

---

Insert Table 3 and Table 4 about here

---

reliability coefficients of widely used standardized tests.

2. Internal consistency (split-half reliability) was assessed by correlating IPDT group test part 1 scores with group test part 2 scores and corrected by the Spearman-Brown Formula to yield reliability coefficients for the test as a whole (Table 4). These coefficients are consistently high and are in the range of consistency coefficients of widely used standardized tests.

3. Total IPDT group test and retest scores were compared by an analysis of variance. A significant phase effect was found, F (1,120) = 6.242, $p < .02$, but a post hoc analysis indicated that the only significant difference in test-retest scores was for group 2, F (1,120) = 16.138, $p < .001$. Figure 2 clearly shows the similarity in means for all groups except two.

Insert Figure 2 about here

These data on reliability correspond closely to reliability studies on other Piaget tests (Burney, 1974; Lawson, 1977; Nelson, 1969; Sheehan, 1970). The lower test-retest correlations in grade 6 can be explained and indeed are actually predicted by Piaget's theory. Students 11-12 years of age are in the transitional stage between concrete and formal operations and their reasoning is not consistent. This is shown by the fact that the variances in scores for grade 6 are larger than any other group (See Table 5). The

group means do not change much from test to retest, but there is

---

Insert Table 5 about here

---

a great deal of individual variability in responses within the group on each test administration.

The significant test-retest difference found in the scores of group 2 is more difficult to explain. Since phase 2 retest scores for grades 1 and 2 are virtually identical, and since group 1 did not change significantly from test to retest, the most likely explanation for the group 2 change is that it is an artifact of the first group test administration. A review of that particular testing session revealed that it was the tester's first test administration and the only session where there were problems maintaining a serious, quiet environment for concentration. The low scores on the first administration for group 2 might well have been due to the students' lack of seriousness in taking the test.

The general conclusions concerning reliability, therefore, are that the IPDT scores are stable over a short period for a wide range of age groups, and that scores are not seriously affected by situational testing variables. Furthermore, the inventory is internally consistent and appears to be measuring a general, lasting trait, rather than a specific, temporary trait.

Validity

To evaluate concurrent validity, i.e. to determine if the

paper-pencil IPDT was assessing cognitive development in the same manner as Piaget's individual interviews and demonstrations, an analysis of variance was performed to determine if there were any significant differences between the individual test and the identical items in the group test. Figure 3 illustrates the finding that there was no significant difference in group test scores

---

Insert Figure 3 about here

---

and individual test scores at any grade level, F (1,148) = .848, $p > .20$. Although the difference for grade three is large, it is not a significant difference.

Construct validity was investigated by comparing mean scores for each group on the IPDT. Additional analyses of variance showed a significant grade effect in both group, F (3,120) = 70.36, $p < .001$, and individual, F (3,152) = 72.518, $p < .001$, IPDT scores; however, a *post hoc* analysis revealed no significant difference in group or individual scores between 9th grade and college, F (1,120) = .83, $p > .20$; F (1,152) = 1.24, $p > .20$. These results would be predicted by Piaget: the 8-9-year-olds are in the concrete stage and should perform differently from 11-12-year-olds, who are in transition, and both groups should be different from 14-15-year-olds who are in the formal stage. Since the 14-15-year-olds and college students should both be in the formal stage, no difference is expected.

The same pattern of increasing mean scores on two different kinds of administration of the IPDT shown in Figures 2 and 3 is also evident in a breakdown of the 18 subtests by problem area illustrated in Figures 4 through 8. The criterion of 75% or more

---

Insert Figures 4 through 8 about here

---

correct responses on items in a subtest (i.e. at least three out of the four items) achieved by 75% or more of a group is used to define mastery of a subtest concept (such a criterion has been used by Elkind, 1962). The data shows 15 out of 18 concepts (83%) have been mastered by college students, 12 out of 18 (67%) by 9th graders, 5 out of 18 (28%) by 6th graders, and none by 3rd graders. Mastery clearly increases with age, and the big increments between grades 3, 6 and 9 are evident.

When IPDT group scores were compared to other standardized instruments, the results showed significant correlations with achievement tests and, with one exception, IQ tests (Table 6). For

---

Insert Table 6 about here

---

college students, the highest correlation with an achievement measure was with high school point average, e.g. for group 11, $\underline{r} = .71$, $\underline{N} = 21$, $\underline{p} < .001$. These data suggest the IPDT is measuring a factor related to both achievement and intelligence, and are

comparable to the findings of others who have compared the results of their Piaget tests to standardized tests (e.g. Lawson, 1977).

The final and least objective indicator of validity was the phase three informal interviews with the public school students. Ten subjects from each grade (groups 3, 6, 9) were randomly selected for interviews. They were shown and asked to explain their answer to the most frequently missed items on the IPDT. The unequivocal finding of both investigators interviewing separately was that students who missed items had not achieved the requisite cognitive structures; neither the paper-pencil, multiple-choice format of the inventory nor the way any individual items were presented appeared to significantly mislead the students into an incorrect answer.

The general conclusion concerning validity is that the IPDT basically shows the developmental progression of reasoning found by Piaget and other researchers in the five major areas included on the inventory, and that it yields a result similar to the traditional, individually administered tasks even though using a paper-pencil format. These indicators, closely corresponding to similar validity studies of other tests (Bart, 1972; Burney, 1974; Lawson, 1977; Renner, 1977), provide initial evidence that the IPDT has concurrent and construct validity.

Other Results

A few other findings are noteworthy. Table 5 summarizes the

results of scores obtained from all students given the group test (excluding groups 1, 4, 7 and 10) and lists the five most difficult subtests for each group. The concepts of classification (classes) and conservation of volume are difficult for all ages, and students who should be in formal operations still have difficulty with perspective (shadows), conservation of length (distance) and laws (rotations). Figures 4 through 8 graphically illustrate this finding: no group mastered classes, distance, or rotations, and only college students mastered conservation volume, images (shadows), and probability--and these were barely mastered. These results clearly support recent studies that show Piaget was incorrect about the timing of the attainment formal thought (e.g. Elkind, 1962; Towler & Wheatley, 1971).

An analysis of scores among full-time college students (Table 7) revealed a slightly higher score for science majors than

---

Insert Table 7 about here

---

for non-science majors, but the difference was not statistically significant. It is interesting to note that the only perfect score on the inventory out of the 542 subjects tested was made by a top science student.

The administration of the IPDT to the two mentally retarded groups resulted in means of 21.5 for the 7-9-year-old group ($N = 8$), and 25 for the 19-45-year-old group ($N = 14$). Because

of the small sample size and because individual IQ's and background information could not be obtained for these subjects, it is difficult to make any interpretations of the results. It would appear, however, that the IPDT could be of use in identifying particular strengths for mildly or moderately retarded adults.

Concerning correlates of scores with subject background information, the comparisons of total IPDT group scores with sex, number of siblings, birth order, or neighborhood showed no significant trends. There clearly is a relationship between age and score, but not within a grade. A correlation of -.50 ($N = 53$, $p < .001$) between IPDT group score and age for subjects over 22 years is consistent with other findings on the decline of formal thought through adulthood (Coleman, 1973).

Total group IPDT scores correlated poorly with introductory college chemistry grades, $r = .28$, $N = 89$, $p < .02$. A higher correlation was found between total IPDT scores and grades in an introductory college logic course, $r = .62$, $N = 20$, $p < .01$. This correlation was approximately the same as that of high school rank and SAT scores with logic grades. These correlations are roughly similar to findings of grade correlations and other Piaget tests (Albanese, et.al. 1976; Good, et.al., 1977).

Usefulness

The IPDT proved to be a quick, easily administered test which can be given to normal subjects 8 years and older in groups and mentally retarded on an individual basis. Since minimal reading

proficiency is required, minority and culturally deprived students with reading problems would not appear to be seriously disadvantaged by the test, although this study does not deal with such samples. Most normal subjects from age 8 through adulthood immediately understood how to take the inventory, and most 8-year-olds had little difficulty reading the items. Most normal subjects required about 45 minutes to complete the inventory; the mentally retarded subjects generally needed more time. In addition, all ages seemed to find the inventory intrinsically interesting and challenging; no one seemed to dislike the test, and most found it enjoyable or even fun. Some adults found some of the items amusing and emitted periodic chuckles.

Several uses are foreseen. Because of the ease and economy of administration, and the inventory's comprehensiveness in covering five problem areas and both concrete and formal tasks (most paper-pencil tests cover one or two problem areas and either concrete or formal task), Piaget researchers could make use of this inventory to further study cognitive development over a wide age range and with different socioeconomic or minority groups. Educators could use the IPDT to aid in curriculum assessment, especially in determining readiness and subsequent ordering for certain science and math courses. The IPDT, used in this manner by Piaget proponents in a school system, could help demonstrate to non-Piagetians the unfolding of specific thought concepts and the importance of designing a curriculum which facilitates reasoning development, for normal as well as exceptional children. In addition, the IPDT

appears to be a promising instrument for use by educators who have developed Piaget-based courses and who need an instrument to identify students with reasoning difficulties, or would like to assess pre-test or posttest performance.

Although there appears to be wide potential for this instrument, clearly a word of caution must be added to prospective users. There are definite limitations to this inventory. As most Piagetians contend, a single score on a paper-pencil inventory cannot yield the same kind of qualitative information that an individual interview can--most especially the reasons for the subject's response. The IPDT therefore would be best used in situations where individual interviews are impractical or impossible, and where a quick, gross measure of functioning is required. Even then, significant decisions on placement of a student certainly should not be made entirely on the strength of the IPDT score. Interpretation of scores should only be made in light of Piaget's theory; the IPDT is not designed to be used as an IQ test or an achievement test.

Users should also keep in mind that the IPDT is directed more at concrete thought than formal thought. As would be expected, therefore, for junior high age and older, there is a ceiling effect on parts of the test since 90% or more of 9th graders got the same 35 items correct, and 90% of college students got the same 40 items correct. It appears that roughly half of the test is too easy for subjects 14-15 years and older; however, with recent evidence of regression of thought in old age (Coleman, 1973; Papalia, Kennedy, & Sheehan, 1973), the concrete operational emphasis of the test would

make this inventory quite useful for life-span studies. So the remaining half of the test does appear to be of value in screening specific concrete and some formal reasoning difficulties in adults, but the inventory is not designed to indicate proficiency in all areas of formal thought.

For use with normal junior and senior high school and college students, a new form of the inventory could be printed to omit the following eight subtests: quantity (#1), levels (#2), sequence (#3), matrix (#5), symbols (#6), movement (#8), seriation (#10), and inferences (#17). Possibly weight (#4), inclusion (#16), and perspective (#7) could also be eliminated. The resulting inventory, while half as long, should be as reliable and valid as the complete IPDT.

Conclusion

In light of the widespread search for standardized instruments to assess Piaget's tasks, the IPDT is clearly a promising theoretical and applied research instrument. The data presented in this study on reliability and validity certainly are a beginning at standardization of the instrument, and provide the necessary, initial background information for judicious use of the IPDT by researchers.

Much more research needs to be done with the instrument before it might be recommended for widespread, applied use. Norms for each age level and various socioeconomic groups need to be determined, and additional reliability and validity studies using a more representative sample need to be done. A more comprehensive analysis of the items also might lead to a revision of the inventory beyond the suggestion on

eliminating certain subtests for older subjects.

This study has only been the beginning of a long path of research to develop a highly effective instrument to assess cognitive development, a needed tool in the empirical investigation of Piaget's theories.

References

Albanese, M. Brooks, D. W., Day, V. W., Koehler, R. A., Lewis, J. D., Marianelli, R. S., Rack, E. P., & Tomlinson-Keasey, C. Piagetian criteria as predictors of success in first year courses. Journal of Chemical Educations, 1976, 53 (9), 571-572.

Arons, A. B. Cultivating the capacity for formal reasoning: Objectives and procedures in an introductory physical science course. American Journal of Physics, 1976, 44 (9), 834-838.

Barnes, G. Scores on a Piaget-type questionnaire versus semester grades for lower-division college physics students. American Journal of Physics, 1977, 45 (9), 841-847.

Bart, W. M. Construction and validation of formal reasoning instruments. Psychological Reports, 1972, 30, 663-670.

Bart, W. M. Issues in measuring formal operational reasoning. Genetic Epistemologist, 1978, 8 (1), 3-4.

Burney, G. The construction and validation of an objective formal reasoning instrument (Doctoral dissertation, University of Northern Colorado, 1974). Dissertation Abstracts International, 1974. (University Microfilms No. 75-5403)

Coleman, S. B. The effect of aging on Piaget's developmental stages: A study of cognitive decline (Doctoral dissertation, Temple University, 1973). Dissertation Abstracts International, 1973, 34 (3A), 1122A-1123A. (University Microfilms No. 73-18,712)

Elkind, D. Quantity conceptions in college students. Journal of Social Psychology, 1962, 57, 459-465.

Elkind, D. Two approaches to intelligence: Piagetian and psychometric. In D. R. Green, M. P. Ford, & G. B. Flamer (Eds.), Measurement and Piaget. New York: McGraw-Hill, 1971.

Filer, A. A. Piagetian cognitive development in normal and in emotionally disturbed children. Unpublished doctoral dissertation, University of Rochester, 1972.

Fuller, R. G., Karplus, R., & Lawson, A. E. Can physics develop reasoning? Physics Today, February, 1977, 23-28.

Furth, H. An inventory of Piaget's developmental tasks. Washington, DC: Center for Research in Thinking and Language, Department of Psychology, Catholic University, 1970.

Goldschmid, M. L., & Bentler, P. M. Manual: Concept assessment kit - conservation. San Diego: Educational and Industrial Testing Service, 1968.

Good, R., Mellon, E., & Kromhout, B. A model for facilitating formal thought in the college science student. Paper presented at the meeting of the American Chemical Society, New Orleans, March 1977.

Herron, J. D. Piaget for chemists. Journal of Chemical Education, 1975, 52 (3), 146-150.

Herron, J. D. More Piaget for chemists: Things that I wish I had told you. Paper presented at the $2YC_3$ Conference, Kansas City, October 1976.

Herron, J. D. Piaget applied: Suggestions for inaction. Paper presented at the American Chemical Society Conference, New Orleans, March 1977.

Herron, J. D. Piaget in the classroom; Guidelines for applications. Journal of Chemical Education, 1978, 55 (3), 165-170.

Inhelder, B., & Piaget, J. The growth of logical thinking from childhood to adolescence. New York: Basic Books, 1958.

Kaufman, A. S. Piaget and Gesell: A psychometric analysis of tests built from their tasks. Child Development, 1971, 42 (5), 1341-1360.

Laurendeau, M. & Pinard, A. Causal thinking in the child. New York: International Universities Press, 1962.

Lawson, A. E. The development and validation of a classroom test of formal reasoning. Paper presented at the annual convention of the National Association for Research in Science Teaching, Cincinnati, March 1977.

Longeot, F. Un essai d'application de la psychologie genetique a la psychologic differentielle. Bulletin de l' Institute National d'Etude, 1962, 18 (3), 153-162.

Lunzer, E. A. Problems of formal reasoning in test situations. In P. H. Mussen (Ed.), European research in cognitive development. Monograph of the Society for Research in Child Development, 1965, 30 (2).

McGrew, J. G., Wonkka, R. E., & Loudin, N. G. Piagetian measure as a predictive device in chemistry courses. Paper presented at the meeting of the American Chemical Society, New Orleans, March 1977.

McKinnon, J. W., & Renner, J. W. Are colleges concerned with intellectual development? American Journal of Physics, 1971, 39, 1047-1052.

Milakofsky, L., & Patterson, H. O. The reliability and validity of an instrument to assess Piaget's tasks. Paper presented at the meeting of the Division of Chemical Education, 174th American Chemical Society National Meeting, Chicago, August 1977.

Milakofsky, L., & Patterson, H. O. Chemical education and Piaget: A new paper-pencil inventory to assess cognitive functioning. Journal of Chemical Education, in press, February 1979.

Nelson, R. J. An investigation of a group test based on Piaget's concepts of number and length conservation and its ability to predict first-grade arithmetic achievement (Doctoral Dissertation, Purdue University, 1969). Dissertation Abstracts International, 1970, 30 (9-A), 3644. (University Microfilms No. 70-5311)

Papalia, D. E., Kennedy, L., & Sheehan, N. Conservation of space in noninstitutionalized old people. The Journal of Psychology, 1973, 84 (1), 75-79.

Patterson, H. O. Cognitive development in the manifest content of children's reported dreams. Unpublished master's thesis, Bucknell University, 1975.

Patterson, H. O., & Milakofsky, L. The reliability and validity of a paper-pencil Piaget inventory. Paper presented at the

Eighth Symposium of the Jean Piaget Society, Philadelphia, May 1978.

Raven, R. J. The development of a test of Piaget's logical operations. Science Education, 1973, 57 (3), 377-385.

Renner, J. W. Evaluating intellectual development using written responses to selected science problems (NSF Report, Grant No. EPP 75-19596, Analysis of cognitive processes). University of Oklahoma, Norman, OK, 1977.

Renner, J. W., & Lawson, A. E. Promoting intellectual development through science teaching. The Physics Teacher, 1973, 11, 273-276.

Ross, B. M. Report on An Inventory of Piaget's Developmental Tasks in testing Navaho Indian Children. Unpublished manuscript, Catholic University, 1968.

Rowell, J. A., & Hoffman, P. J. Group tests for distinguishing formal from concrete thinkers. Journal of Research in Science Teaching, 1975, 12 (2), 157-164.

Shayer, M., & Wharry, D. Piaget in the classroom I: Testing a whole class at the same time. Unpublished manuscript, Chelsa College, University of London, 1975.

Sheehan, D. J. The effectiveness of concrete and formal instructional procedures with concrete - and formal - operational students (Doctoral dissertation, State University of New York at Albany, 1970). Dissertation Abstracts International, 1970, 31 (6A), 2748A. (University Microfilms No. 70-25,479)

Sills, T. W. Developmental and evaluative study of a written Piagetian paper-and-pencil test of formal operations. Unpublished doctoral dissertation, Purdue University, 1977.

Tisher, R. P. A Piagetian questionnaire applied to pupils in a secondary school. Child Development, 1971, 42 (5), 1633-1636.

Tomlinson-Keasey, C. Introduction to formal operation tasks. Unpublished manuscript, University of Nebraska, 1975.

Towler, J. O., & Wheatley, G. Conservation concepts in college students: A replication and critique. Journal of Genetic Psychology, 1971, 118, 265-270.

Tuddenham, R. D. Theoretical regularities and individual idiosyncrasies. In D. R. Green, M. P. Ford, & G. B. Flamer (Eds.), Measurement and Piaget. New York: McGraw-Hill, 1971.

Footnotes

Paper presented to the Division of Educational Psychology at the Eighty-Sixth Annual Convention of the American Psychological Association, Toranto, Ontario, Canada, August 28, 1978.

Requests for copies of this paper or copies of the authors' other papers on this subject should be sent to Henry O. Patterson and Louis Milakofsky, The Pennsylvania State University, Berks Campus, R.D.#5, Tulpehocken Road, P.O. Box 2150, Reading, Pennsylvania, 19608.

Parts of this study were funded under a grant from The Pennsylvania State University Commonwealth Campus Scholarly Activity Fund (Phase III), The Berks Campus Scholarly Activity Fund, and the College of Science and College of Liberal Arts Computer Funds.

The authors gratefully acknowledge the generous cooperation and assistance of the following: Dr. Stanley T. Dubelle, Mr. Christ J. Bucolo, Mr. Stephen Gancar, and Mr. Calvin Bossler of the Governor Mifflin School District; Dr. Robert E. Brown, Berks Campus; Mr. Scott Schaeffer, Miss Cathy Plushanski, Mr. William Hockley, Mr. James Henry, Mr. Joel Cleveland, Miss Robin Juckem, Mr. Thomas Gerhard, and Miss Nancy Miller, student assistants; and Mrs. Nancy Patterson, Miss Barbara Stutzman, and Miss Sandy Kreis.

[1]Printed copies of this inventory can be purchased for $6 each from Dr. Bruce M. Ross, Department of Psychology, Catholic University, Washington, DC 20017.

[2]The IPDT booklets and answer key purchased from Catholic University for use in this study contained two errors which were corrected in this study for all administrations: (a) the correct answer for item 11 is "D", not "B" as indicated on the answer key; (b) for item 71, in order for answer "D" to be correct, as indicated on the answer key, the "D" figure would have to be changed; we chose to place dots in the third ball, leaving the first two white.

Table 1

Reliability-Validity Research Design

Subjects		Phase I Testing		Phase II Testing	Phase III Testing
Grade 3 (8-9-Yr.-Olds) N = 58	Grp. 1	Indiv.	Grp.	Retest Grp.	
	Grp. 2	Grp.	Indiv.	Retest Grp.	
	Grp. 3		Grp.Pt.1	Grp.Pt.2	Individual Interviews
Grade 6 (11-12-Yr.-Olds) N = 57	Grp. 4	Indiv.	Grp.	Retest Grp.	
	Grp. 5	Grp.	Indiv.	Retest Grp.	
	Grp. 6		Grp.Pt.1	Grp.Pt.2	Individual Interviews
Grade 9 (14-15-Yr.-Olds) N = 60	Grp. 7	Indiv.	Grp.	Retest Grp.	
	Grp. 8	Grp.	Indiv.	Retest Grp.	
	Grp. 9		Grp.Pt.1	Grp.Pt. 2	Individual Interviews
College (18-19-Yr.-Olds) N = 62	Grp.10	Indiv.	Grp.	Retest Grp.	
	Grp.11	Grp.	Indiv.	Retest Grp.	
	Grp.12		Grp.Pt.1	Grp.Pt.2	

Table 2

Content Summary of IPDT

Subtest No.	Subtest Name	Problem Area	Concept Assessed	Approx. Mastery Age
1	Quantity	Conservation	Conservation of Quantity	7-8
2	Levels	Images	Transformational Imagery	9-10
3	Sequence	Relations	Ordinal Relations	7-8
4	Weight	Conservation	Conservation of Weight	9-10
5	Matrix	Classification	Classification	7
6	Symbols	Classification	Combinativity	7-8
7	Perspective	Images	Perspective	9-10
8	Movement	Images	Kinetic Imagery	8
9	Volume	Conservation	Conservation of Volume	11-12
10	Seriation	Relations	Ordinal Relations	7-8
11	Rotation	Laws	Kinetic Imagery	8-9
12	Angles	Laws	Reciprocal Implication	12
13	Shadows	Images	Perspective	9-10
14	Classes	Classification	Classification	12-13
15	Distance	Conservation	Conservation of Length	10
16	Inclusion	Classification	Verbal Class Inclusion	11-12
17	Inference	Relations	Verbal Transitivity	11-12
18	Probability	Laws	Probability	10-11

Table 3

Pearson Product-Moment Correlation

Coefficients for IPDT Group Test-Retest Scores

	Grade							
	3		6		9		College	
Group	1	2	4	5	7	8	10	11
r	.87***	.75***	.40	.62**	.84***	.86***	.95***	.67***
n	20	20	20	19	20	20	21	21

*$p < .05$.

**$p < .01$.

***$p < .001$.

Table 4

Split-Half Reliability Coefficients for IPDT

	Grade			
	3	6	9	College
Group	3	6	9	12
n	17	18	17	20
r^a	.71	.63	.84	.71

[a]Corrected by Spearman-Brown Formula.

Table 5

Mean Scores of IPDT by Grade with Mean Scores of Five Most Difficult Subtests

Grade	IPDT[a]			Subtest[b]			
	$\bar{x}$	SD	n	No.	Name	$\bar{x}$	SD
3 (8-9-Yr.-Olds)	30.05	5.75	20	9	Volume	.60	.68
				7	Perspective	.75	.91
				14	Classes	.90	.72
				11	Rotation	1.00	.65
				13[c]	Shadows	1.45	1.19
				18[c]	Probability	1.45	1.15
6 (11-12-Yr.-Olds)	47.31	10.22	94[d]	11	Rotation	1.70	1.21
				14	Classes	1.73	1.13
				18	Probability	1.91	1.22
				9	Volume	1.98	1.20
				15	Distance	2.18	.81
9 (14-15-Yr.-Olds)	57.00	6.84	20	14	Classes	1.95	1.39
				15	Distance	2.35	.81
				13	Shadows	2.55	1.10
				9	Volume	2.65	1.18
				18	Probability	2.80	1.24
College (17+Yr.-Olds)	62.27	5.92	226[d]	14	Classes	2.27	1.37
				15	Distance	2.74	.76
				11	Rotation	2.96	1.07
				13	Shadows	3.06	.84
				9	Volume	3.25	1.03

[a] Maximum score = 72.

[b] Maximum score = 4.

[c] Tie score.

[d] Includes Ss who were not part of the reliability study.

Table 6

Pearson Product-Moment Correlation Coefficients for Group IPDT Scores and Standardized Test Scores

Test Score	Grade 3 Grp. 2	Grade 6 Grp. 5	Grade 9 Grp. 8	College Grp. 11	Adults ≥22 Yrs.
n	20	19	20	21	53
Full Scale IQ[a]	.42	.49*	.68***		
Total SAT				.44*	.63***
Composite Achievement[b]	.58**	.87***	.46*		

[a]Lorge-Thorndike Intelligence Test.

[b]Iowa Tests of Basic Skills.

*$p < .05$.

**$p < .01$.

***$p < .001$.

Table 7

Mean Subtest Scores for College Science-Engineering and Non-Science Students

Subtest Number	Subtest Name	Mean Score		F*
		Science[a]	Non-Science[b]	
1	Quantity	3.42	3.29	.707
2	Levels	3.89	3.67	2.157
3	Sequence	3.85	4.00	1.855
4	Weight	3.97	3.86	2.308
5	Matrix	3.89	3.95	.388
6	Symbols	3.70	3.81	.834
7	Perspective	3.82	3.57	3.132
8	Movement	3.70	3.62	.252
9	Volume	3.36	3.10	1.184
10	Seriation	3.89	3.95	.486
11	Rotation	3.08	2.90	.532
12	Angles	3.62	3.33	2.411
13	Shadows	3.17	2.95	1.605
14	Classes	2.44	2.19	.726
15	Distance	2.78	2.52	1.474
16	Inclusion	3.71	3.48	1.414
17	Inferences	3.82	3.86	.079
18	Probability	3.38	3.19	.685
TOTAL		63.5[c]	61.2[c]	3.056

Note. Maximum Score = 4.

[a] $n = 89$.

[b] $n = 21$.

[c] Maximum Score = 72.

*p n.s. with df of 1,108.

Figure Captions

Figure 1. Sample page from the IPDT showing the example and the last item on the conservation of volume subtest.

Figure 2. Mean scores of IPDT group tests and retests one month later.

Figure 3. Mean scores of identical items on group test part 1 and individual test for each grade.

Figure 4. Percentage of subjects by grade attaining at least three out of four items correct on each of four classification subtests.

Figure 5. Percentage of subjects by grade attaining at least three out of four items correct on each of four conservation subtests.

Figure 6. Percentage of subjects by grade attaining at least three out of four items correct on each of four images subtests.

Figure 7. Percentage of subjects by grade attaining at least three out of four items correct on each of three laws subtests.

Figure 8. Percentage of subjects by grade attaining at least three out of four items correct on each of three relations subtests.

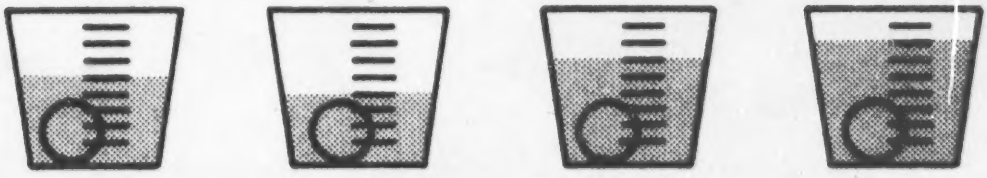
example

Here is a clay ball.

Here is a jar of water.

We put the ball in the water.

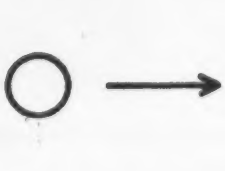
What happens?

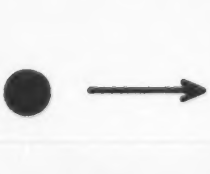


(A) B C D

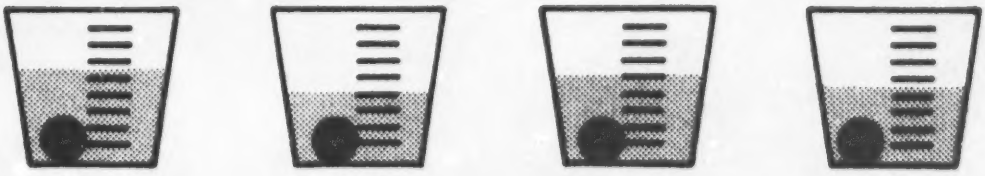
Here is a light ball.

Here is a small, heavy ball.

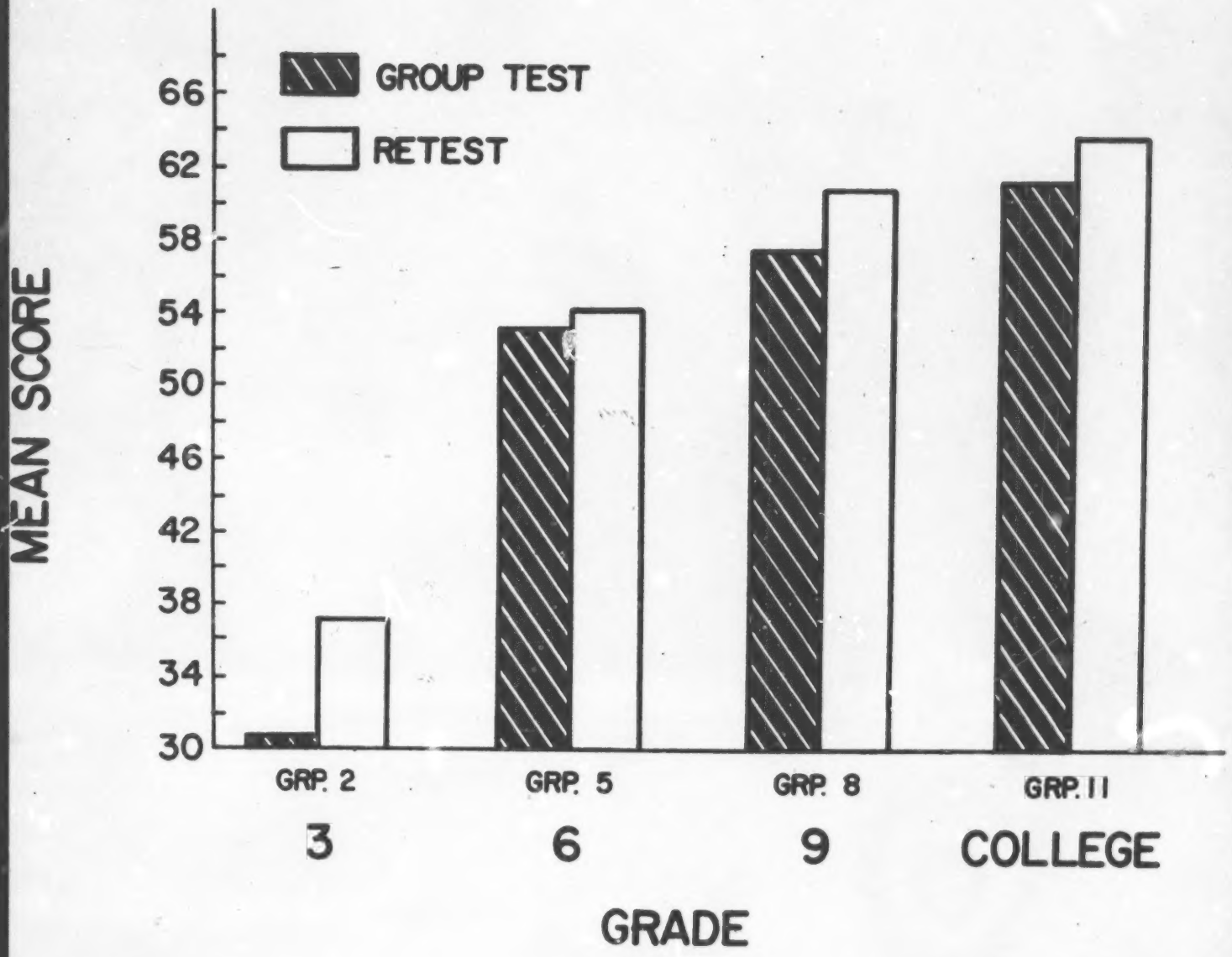


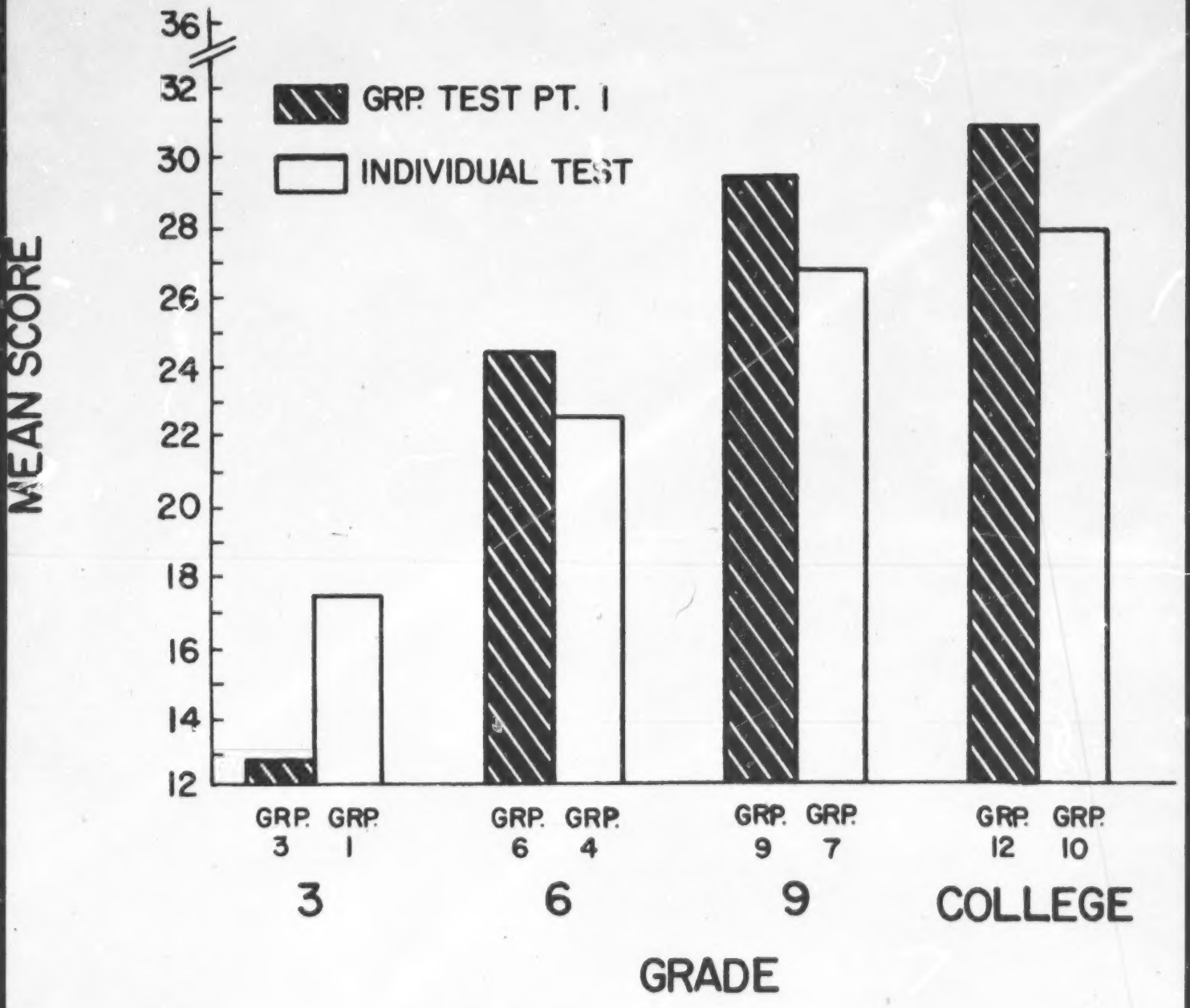


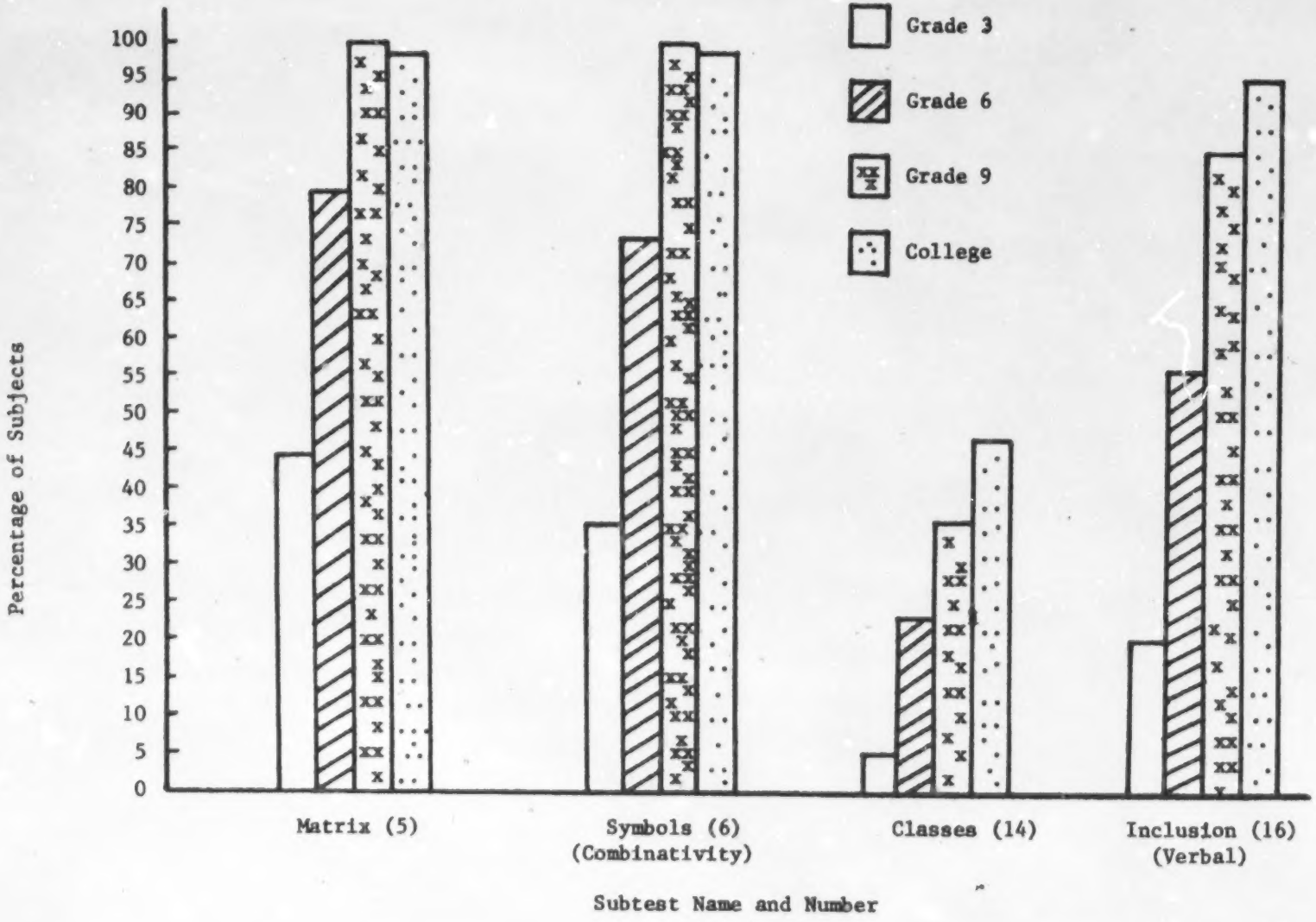
What happens with the small, heavy ball?

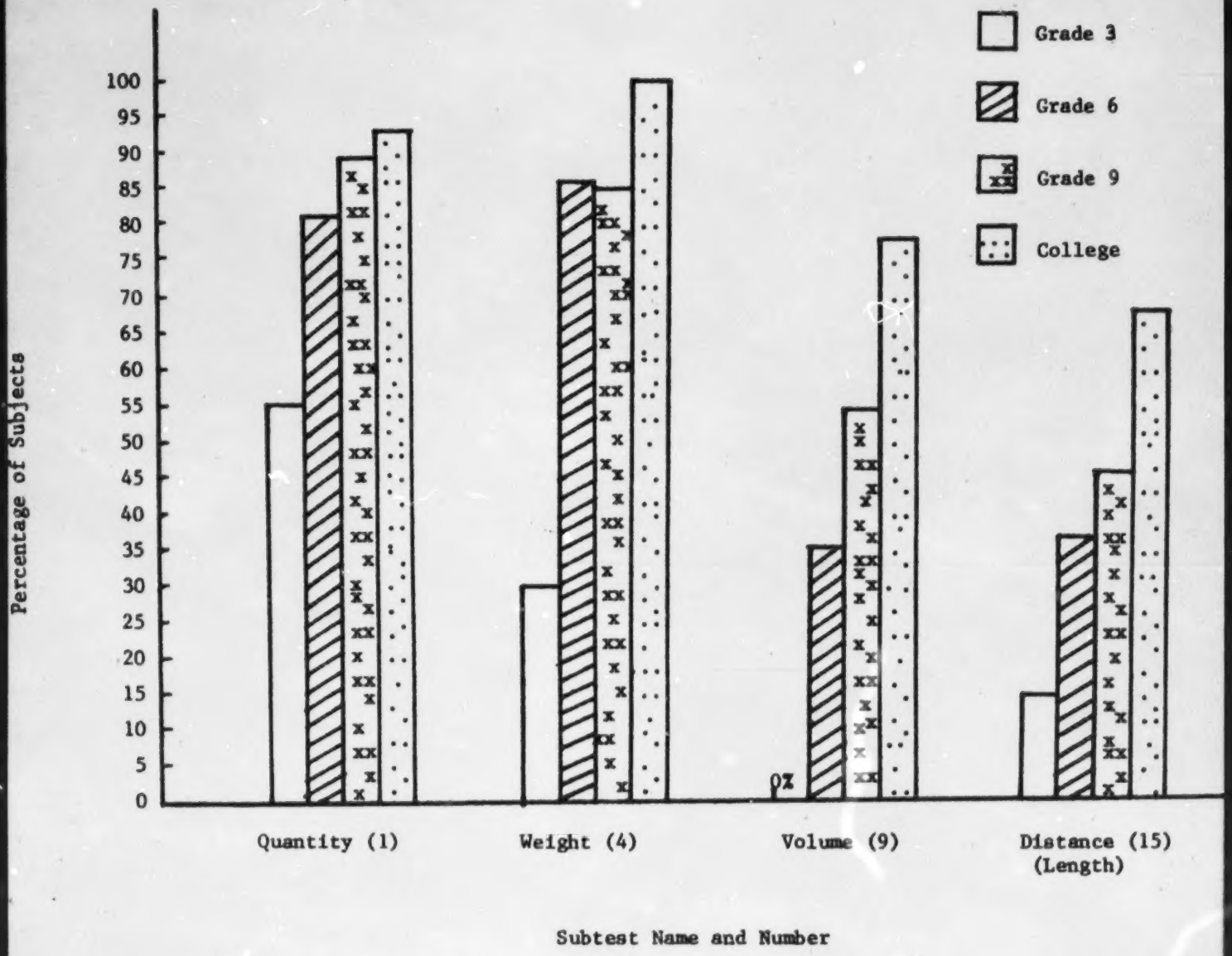


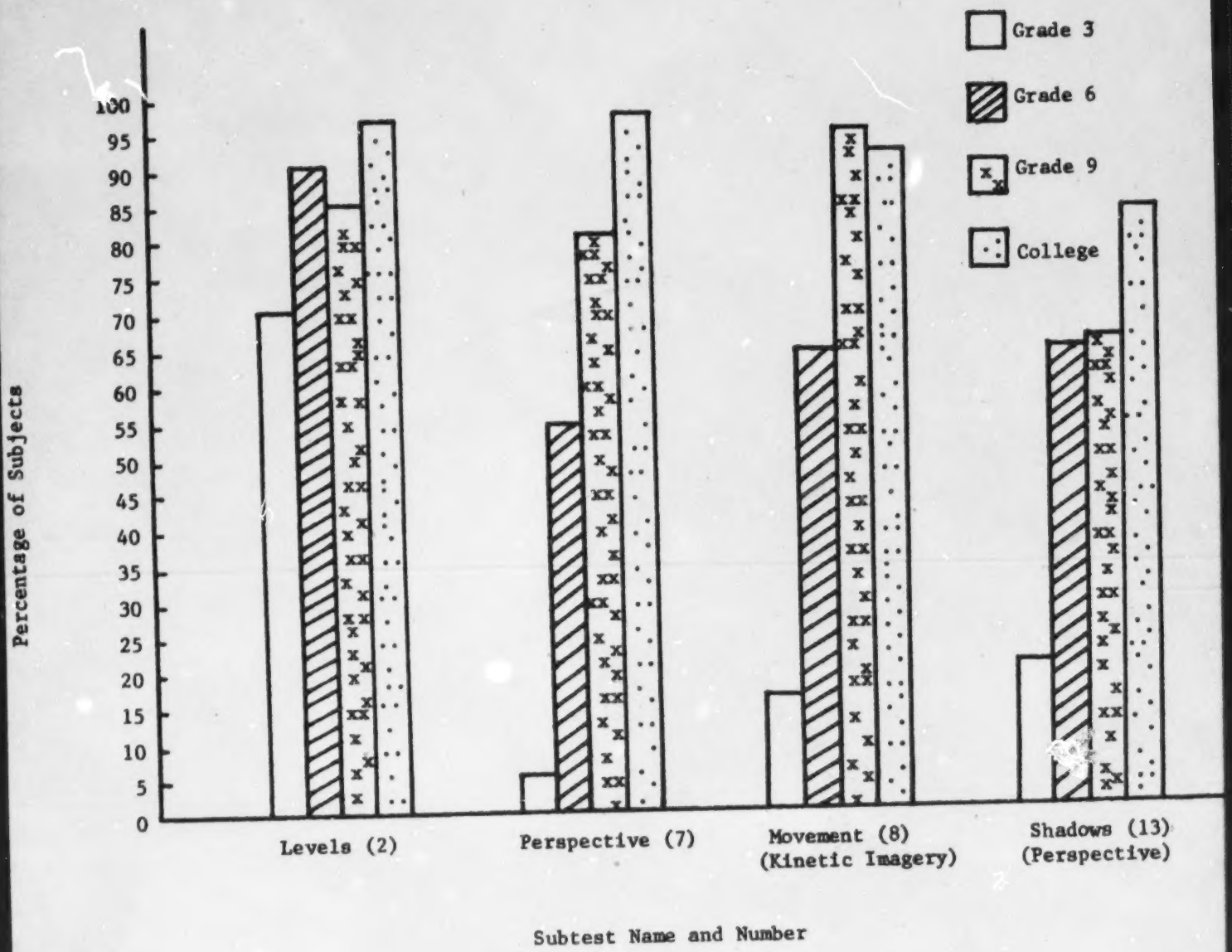
A B C D

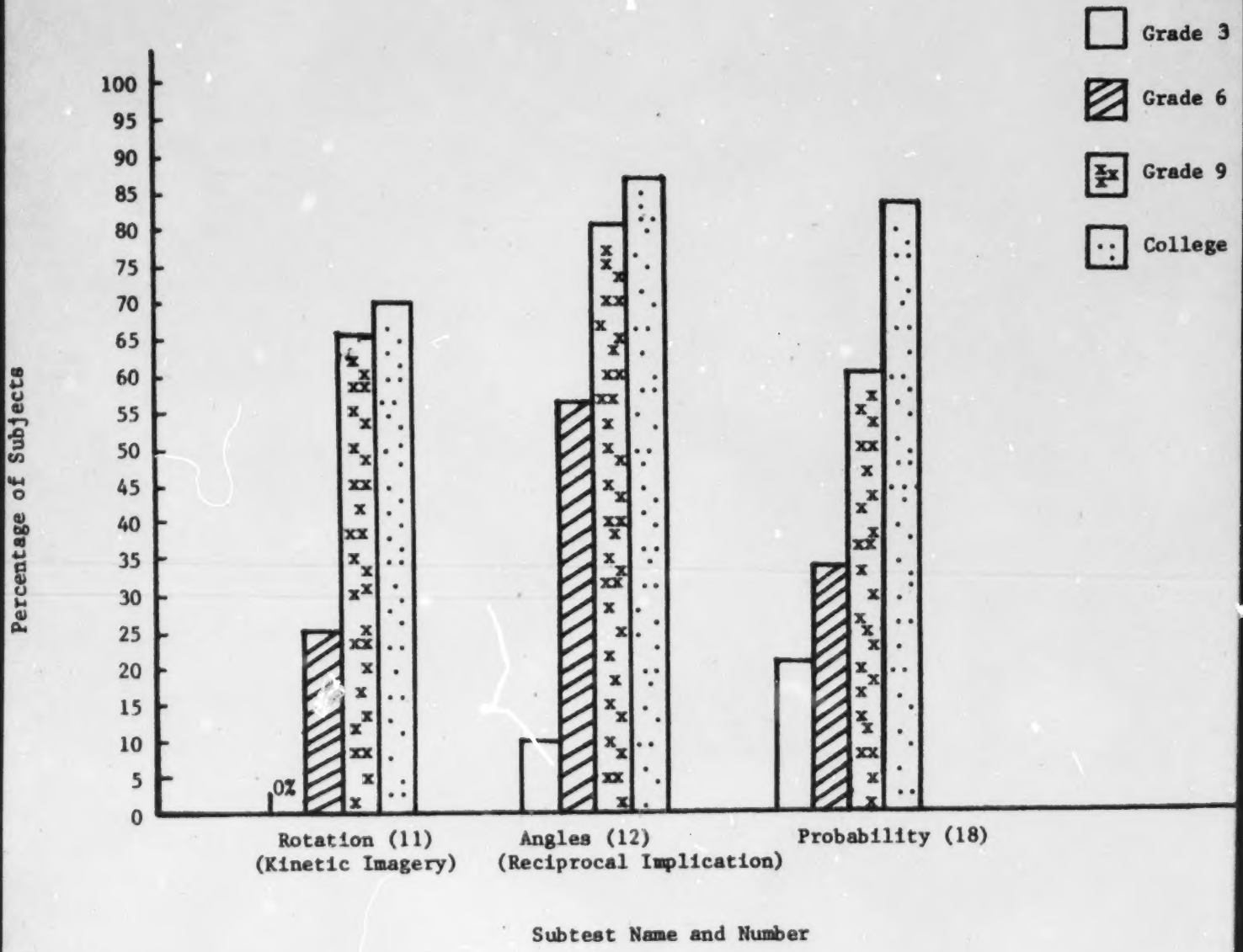

GROUP TEST
RETEST
MEAN SCORE
66
62
58
54
50
46
42
38
34
30
GRP. 2
GRP. 5
GRP. 8
GRP. 11
3
6
9
COLLEGE
GRADE

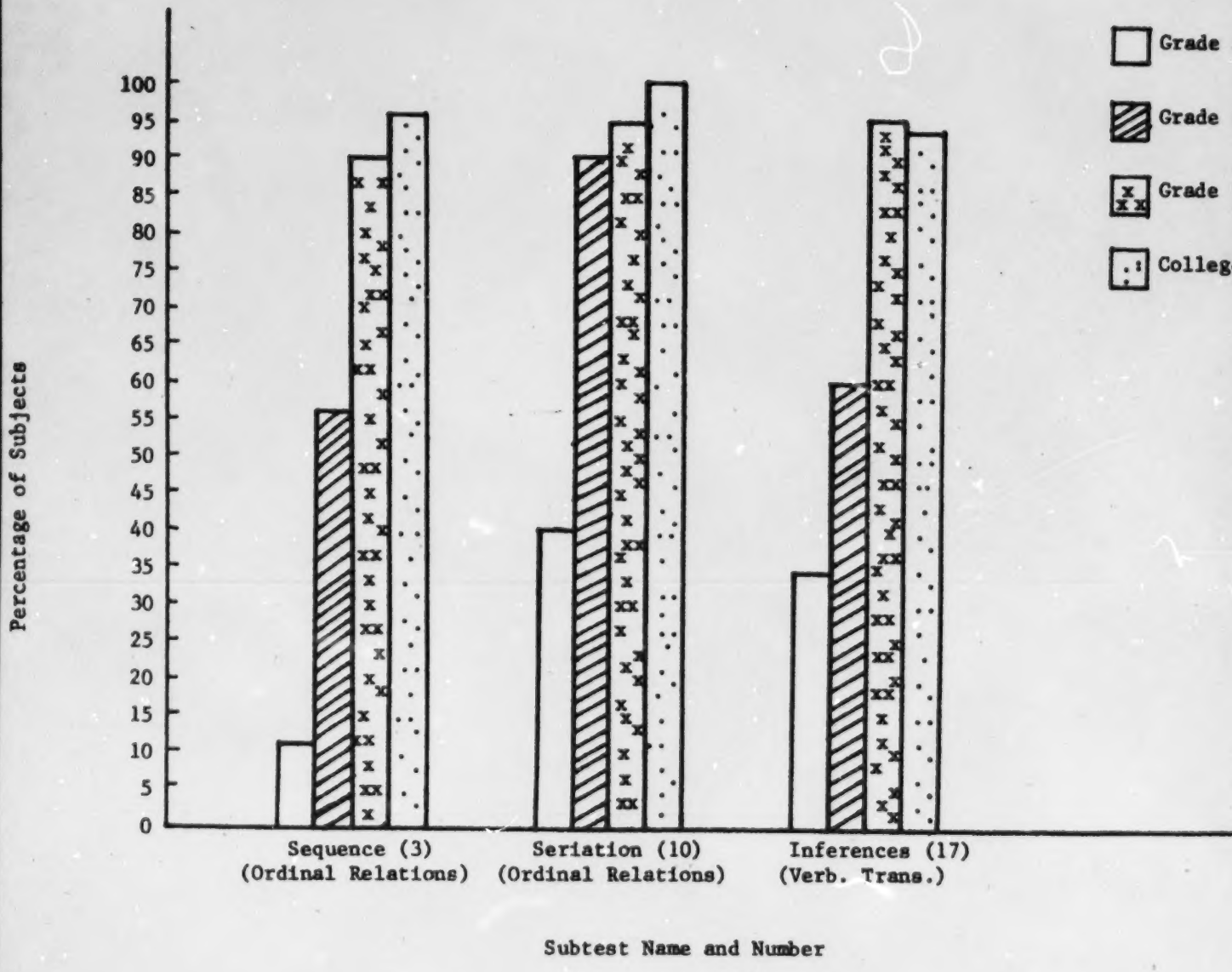

MEAN SCORE
36
32
30
28
26
24
22
20
18
16
14
12
GRP. TEST PT. 1
INDIVIDUAL TEST
GRP. 3
GRP. 1
GRP. 6
GRP. 4
GRP. 9
GRP. 7
GRP. 12
GRP. 10
3
6
9
COLLEGE
GRADE


Grade 3
Grade 6
Grade 9
College
Percentage of Subjects
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
Matrix (5)
Symbols (6)
(Combinativity)
Classes (14)
Inclusion (16)
(Verbal)
Subtest Name and Number


Grade 3
Grade 6
Grade 9
College
Percentage of Subjects
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
0%
Quantity (1)
Weight (4)
Volume (9)
Distance (15)
(Length)
Subtest Name and Number


Grade 3
Grade 6
Grade 9
College
Percentage of Subjects
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
Levels (2)
Perspective (7)
Movement (8)
(Kinetic Imagery)
Shadows (13)
(Perspective)
Subtest Name and Number


Grade 3
Grade 6
Grade 9
College
Percentage of Subjects
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
0%
Rotation (11)
(Kinetic Imagery)
Angles (12)
(Reciprocal Implication)
Probability (18)
Subtest Name and Number


Grade
Grade
Grade
Colleg
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
Percentage of Subjects
Sequence (3)
(Ordinal Relations)
Seriation (10)
(Ordinal Relations)
Inferences (17)
(Verb. Trans.)
Subtest Name and Number